TITLE OF THE INVENTION

ENZYME CATALYZED ISOTOPE LABELING

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of U.S. Provisional Application No. 60/450,034, filed February 26, 2003, the subject matter of which is incorporated herein by reference in its entirety.

BACKGROUND OF THE INVENTION

(1) Field of the Invention

[0002] The invention relates to isotope labeling of peptides and more particularly to enzyme proteolysis of proteins followed by enzyme-catalyzed isotope labeling of peptides.

(2) Description of related art

[0003] It is of interest to introduce stable isotope labels into peptides for a variety of applications. Foremost among these are comparative proteomics studies, where mass spectrometry is used to quantitate the relative amounts of peptides and proteins in two protein pools, e.g., from cell lines with two different health states or histories.

[0004] Other applications include preparation of isotope-labeled internal standards, facilitation of sequencing based on fragment ions in mass spectra, preparation of tracer compounds for chemical and biochemical studies.

[0005] A number of chemical reactions have been proposed, which label the parent protein, and are combined with chemical or enzymatic digestion to produce labeled peptides {(a) Gygi, S.P.; Rist, B.; Gerber, S.A.; Turecek, F.; Gelb, M.H.; Aebersold, R. Nature Biotechnol. 1999, 17, 994. (b) Munchbach, M.; Quadroni, M.; Miotto, G.; James, P. Anal. Chem. 2000, 72, 4047. (c) Geng, M.; Ji, J.; Regnier, F. E. J. Chromatogr., A 2000, 870, 295. (c) Goshe, M. B.; Conrads, T. P.; Panisko, E. A.; Angell, N. H.; Veenstra, T. D.; Smith, R. D. Anal. Chem. 2001, 73, 2578.}.

[0006] Other strategies involve the growth of organisms or cells in isotope labeled media, which lead to incorporation of isotope labels into cellular proteins. These are then isolated and cleaved into labeled peptides { Oda, Y.; Huang, K.; Croos, F. R.; Cowburn, D.; Chait, B. T. *Proc. Natl. Acad. Sci. U.S.A.* 1999, *96*, 6591. (b) Veenstra, T.D.; Martinović, S.; Anderson, G.A.; Paša-Tolić, L.; Smith, R.D. *J. Am. Soc. Mass Spectrom.* 2000, *11*, 78-82. (c) Washburn, M.P., Ulaszek, R., Deciu, C., Schieltz, D.M.; Yates, J.R. III. *Anal. Chem.* 2002, *74*, 1650.}.

[0007] Previously reported {see (a) Yao, X.; Freas, A.; Ramirez, J.; Demirev, P.A.; Fenselau, C. *Anal. Chem.* 2001, *73*, 2836. (b) Reynolds, K.; Yao, X.; Fenselau, C. *J. Proteome Res.* 2002, *1*, 27.} and others { (a) Schnolzer, M.; Jedrzejewski, P.; Lehmann, W. D. *Electrophoresis* 1996, *17*, 945. (b) Wang, Y. K.; Ma, Z.; Quinn, D. F.; Fu, E.W. Anal. Chem. 73, 2001, 3742. (c) Mirgorodskaya, O.A.; Kozmin, Y.P.; Titov, M.I.; Korner, R.; Sonksen, C.P.; Roepstorff, P. *Rapid Commun. Mass Spectrom.* 2000, *14*, 1226.} has been use of serine proteases (trypsin, Lys-C endoprotease and Glu-C endoprotease) to cleave proteins, and simultaneously to incorporate two isotopes of O-18 from highly enriched O-18 water. Two isotopes are incorporated into every new

carboxyl terminus formed, in enzyme catalysis-based cleavage and exchange reactions. The incorporation can exceed 95% if the water in which the proteolysis is carried out is enriched >95% in O-18. A mechanistic scheme is shown in Figure 1, in which one O-18 atom is introduced when the solvent disrupts the covalent intermediate formed between the serine protease and peptide substrates. The second atom of O-18 is introduced when the enzyme rebinds the newly formed peptides and an enzyme-catalyzed oxygen exchange reaction occurs. The high level of enrichment provides a four Dalton increment between peptide pairs representing the two protein pools, which is sufficient for quantitative determinations of isotope ratios by mass spectrometers with appropriate resolution.

[0008] An equation has been published that considers the abundances of natural isotopes and provides the ratio of labeled and unlabeled peptides {see Yao, X.; Freas, A.; Ramirez, J.; Demirev, P.A.; Fenselau, C. *Anal. Chem.* 2001, *73*, 2836, which is hereby incorporated by reference in its entirety.}. It has also been determined the analytical figures of merit to demonstrate the robustness of this approach { Reynolds, K.; Yao, X.; Fenselau, C. *J. Proteome Res.* 2002, *1*, 27.}.

[0009] In most reports {see (a) Yao, X.; Freas, A.; Ramirez, J.; Demirev, P.A.; Fenselau, C. Anal. Chem. 2001, 73, 2836. (b) Reynolds, K.; Yao, X.; Fenselau, C. J. Proteome Res. 2002, 1, 27. (c) Wang, Y. K.; Ma, Z.; Quinn, D. F.; Fu, E.W. *Anal. Chem.* 73, 2001, *3742*. (d) Mirgorodskaya, O.A.; Kozmin, Y.P.; Titov, M.I.; Korner, R.; Sonksen, C.P.; Roepstorff, P. *Rapid Commun. Mass Spectrom.* 2000, *14*, 1226.}, the production of O-18 coded peptides was intended to facilitate comparative proteomic measurements of protein levels in two different protein pools. The production of labeled peptides by

proteolysis reactions requires relatively large amounts of O-18 water. Conditions for the proteolysis reactions must be adjusted to keep the protein substrates in solution, and are not always compatible with optimal enzyme activity. Most critically, the protein substrates have to be dehydrated in order to provide aqueous solutions >95% enriched in O-18 water. Often it is difficult to completely resolubilize the proteins.

[0010] The known methods require chemical synthesis of isotope labeled alkylating agents and chemical reactions to attach these to peptides. Other methods introduce O-18 as a by product of proteolysis reactions, which must be optimized for proteolysis, not for isotope incorporation.

BRIEF SUMMARY OF THE INVENTION

[0011] In this invention, isotope labels are introduced into carboxylic acid groups in peptides under enzyme catalysis. ^{18}O atoms, which are stable to chemical exchange, and can be introduced into all peptides that end in an amino acid residue that is bound by one of a variety of proteolytic enzymes. The enzyme-catalyzed isotope incorporation is conducted in a separate step from enzymatic proteolysis of proteins, under conditions that can be optimized for rapid labeling, minimal enzymatic autolysis, using a minimum of the isotope source.

[0012] Some embodiments of this invention are directed to a method of labeling peptides, comprising the steps of: a) obtaining peptides formed from proteins digested by proteolyic enzymes selected from trypsin, chymotrypsin, Lys-C endoprotease, Glu-C endoprotease, an endoprotease that cleaves at the C-terminal side of an amino acid residue on a protein, or a mixture of two or more of these

proteolytic enzymes; and thereafter b) incorporating isotopic atoms into said peptides in the catalytic presence of proteolyic enzymes selected from trypsin, chymotrypsin, Lys-C endoprotease, Glu-C endoprotease, an endoprotease that cleaves at the C-terminal side of an amino acid residue on a protein, or a mixture of two or more of these proteolytic enzymes.

[0013] In some embodiments of the method of this invention, the proteolytic enzymes or mixture of enzymes in step a) is the same as the proteolytic enzymes or mixture of proteolytic enzymes in step b). In some embodiments, the proteolytic enzymes in step a) and in step b) are trypsin. In some other embodiments, the proteolytic enzymes in step a) and in step b) are a mixture of Glu-C and trypsin.

[0014] In some embodiments of this invention, the proteolytic enzymes or mixture of enzymes in step a) are different from the proteolytic enzymes or mixture of proteolytic enzymes in step b). In some of these embodiments, the proteolytic enzymes in step a) are Lys-C and in step b) are trypsin.

[0015] The proteolytic enzymes in step a) and in step b) can be obtained from any species or from thermophilic microorganisms.

[0016] In this invention, one or two isotopic atoms can be incorporated into a peptide and, in some embodiments, the isotopic atoms are ^{18}O atoms.

[0017] In some embodiments of this invention, the step of incorporating isotopic atoms into said peptides is conducted in $H_2{}^{18}O$.

[0018] In some embodiments of this invention, the method can further comprise mixing labeled peptides obtained from the step of incorporating of isotopic atoms with peptides that have not been labeled to obtain a mix of labeled and unlabeled peptides.

In some of these embodiments, the method can further comprise analyzing the mix of labeled and unlabeled peptides by mass spectrometry. In some of these embodiments, the a ratio of labeled peptides to unlabeled peptides can be calculated with formula (I) or (II)

$$\text{ratio}= \{I_4- (M_4/M_0))\times I_0-(M_2/M_0) \times [I_2-(M_2/M_0) \times I_0] + [I_2 -(M_2/M_0) \times I_0]\}/I_0 \quad (I)$$

$$\text{ratio}= \{I_4- (M_4/M_0))\times I_0-(M_2/M_0) \times [I_2-(M_2/M_0) \times I_0] + 1/2[I_2 -(M_2/M_0) \times I_0]\}/I_0 \quad (II);$$

where I_0, I_2 and I_4 are observed peak areas for a monoisotopic peak for peptides without ^{18}O label, a peak 2Da higher and peak 4Da higher, and M_0, M_2 and M_4 are the theoretical peak areas for the monoisotopic peak for a peptide with a known composition, a peak 2Da higher and peak 4Da higher, respectively.

[0019] In some embodiments of this invention, mass spectrometry utilized can be nanospray, electrospray, LC-MS, LC-MS-MS or matrix-assisted IR or UV laser desorption ionization, high vacuum, atmospheric or low pressure, on a quadropole, quadropole ion trap, time-of-flight, ion cyclotron resonance, magnetic sector ion analyzer or a combination thereof.

[0020] In some embodiments of this invention, the step of incorporating isotopic atoms results in peptides in the catalytic presence of proteolytic enzymes being labeled greater than 90% with one or two ^{18}O atoms. In some of these embodiments, the step of incorporating isotopic atoms results in peptides in the catalytic presence of proteolytic enzymes being labeled greater than 95% with one or two ^{18}O atoms.

[0021] In some embodiments of this invention, the method further comprises a step of storing the peptides overnight after step a) and before step b).

[0022] Some embodiments of this invention are directed to a method of labeling peptides, comprising the steps of: digesting proteins with proteolytic enzymes in an $H_2{}^{16}O$ environment to obtain peptide fragments of said proteins, where said proteolyic enzymes are selected from trypsin, chymotrypsin, Lys-C endoprotease, Glu-C endoprotease, an endoprotease that cleaves at the C-terminal side of an amino acid residue on a protein, or a mixture of two or more of these proteolytic enzymes; and then labeling at least a first portion of said peptide fragments by incorporating ^{18}O atoms into said at least first portion of said peptide fragments in an $H_2{}^{18}O$ environment and in the catalytic presence of proteolyic enzymes selected from trypsin, chymotrypsin, Lys-C endoprotease, Glu-C endoprotease, an endoprotease that cleaves at the C-terminal side of an amino acid residue on a protein, or a mixture of two or more of these proteolytic enzymes.

[0023] In some embodiments of this invention, the method further comprises a step of storing the peptides overnight after the step of digesting proteins and before the step of labeling. In some of these embodiments, the method further comprises the step of: mixing a second unlabeled portion of peptide fragments with the labeled first portion of peptide fragments.

[0024] Some embodiments of this invention are directed to a method of quantitatively analyzing proteins, comprising the steps of: obtaining a protein sample; then dissolving the protein sample in a solution of buffer or chaotropic agent; then forming a peptide sample by digesting the protein sample by adding one or more of trypsin, chymotrypsin, Lys-C endoprotease, Glu-C endoprotease, and/or an endoprotease that cleaves at the C-terminal side of an amino acid residue on a protein; then incubating the peptide

sample at least overnight; then retrieving the peptide sample; then contacting at least a first portion of the peptide sample with said one or more of trypsin, chymotrypsin, Lys-C endoprotease, Glu-C endoprotease, and/or an endoprotease that cleaves at the C-terminal side of an amino acid residue on a protein and with $H_2{}^{18}O$ to label the at least first portion of the peptide sample with ^{18}O. In some of these embodiments, the method further comprises the steps of preparing a second unlabeled portion of said peptide samples by contacting said second portion of the peptide sample with said one or more of trypsin, chymotrypsin, Lys-C endoprotease, Glu-C endoprotease, and/or an endoprotease that cleaves at the C-terminal side of an amino acid residue on a protein and with $H_2{}^{16}O$; and then mixing a second unlabeled portion of said peptide samples with the labeled first portion of the peptide sample; then calculating a ratio of labeled peptides to unlabeled peptides with formula (I) or (II)

$$\text{ratio}= \{I_4- (M_4/M_0))\times I_0-(M_2/M_0) \times [I_2-(M_2/M_0) \times I_0] + [I_2 -(M_2/M_0) \times I_0]\}/I_0 \qquad (I)$$

$$\text{ratio}= \{I_4- (M_4/M_0))\times I_0-(M_2/M_0) \times [I_2-(M_2/M_0) \times I_0] + 1/2[I_2 -(M_2/M_0) \times I_0]\}/I_0 \qquad (II)$$

where I_0, I_2 and I_4 are observed peak areas for a monoisotopic peak for peptides without ^{18}O label, a peak 2Da higher and peak 4Da higher, respectively, and M_0, M_2 and M_4 are the theoretical peak areas for the monoisotopic peak for a peptide with a known composition, a peak 2Da higher and peak 4Da higher, respectively.

BRIEF DESCRIPTION OF THE SEVERAL VIEWS OF THE DRAWINGS

[0025] Embodiments of this invention will now be described in detail with reference to the attached Figures, in which:

[0026] Figure 1 shows the mechanism of double ^{18}O incorporation in peptides;

[0027] Figure 2 shows the dissection of incorporation of two stable isotopes during proteolysis;

[0028] Figure 3 shows MALDI-FT-ICR spectra of YGGFMK at 0% and 10% conversion of the $^{16}O_2$-peptide. Two spectra were normalized according to a total concentration of $^{16}O_2$-peptide, $^{16}O^{18}O$-peptide, and $^{18}O_2$-peptide;

[0029] Figure 4 shows a double reciprocal plot of $1/v_i$ vs $1/[R_o]$ for trypsin-catalyzed ^{16}O-to-^{18}O exchange reaction of YGGFMR;

[0030] Figure 5 shows sequential exchange of two carboxyl oxygens by trypsin catalysis;

[0031] Figure 6 shows the simultaneous determination of pseudo first-order decay of multiple truncated peptides ($^{16}O_2$). They are YGGFMK($^{16}O_2$), YGGFLK($^{16}O_2$), and five $^{16}O_2$-peptides from apomyoglobin. R's and K's in brackets represent the peptide P_1 residues. Rate constants (k_{cat}/K_M's) in the inset were calculated, using k_{cat}/K_M of YGGFMK as an internal standard. Data were based on a single experiment;

[0032] Figure 7 shows sequential exchange of two carboxyl oxygens by serine protease: incorporation of the first oxygen isotope;

[0033] Figure 8 shows MALDI FT-ICR spectra of peptides $^{18}O_2$-encoded by chymotrypsin. The last five amino acid residues to the carboxyl termini for three chymotryptic peptides are shown. Arrows point to would-be positions for the peptides with ($^{16}O_2$);

[0034] Figure 9 shows a protocol for digestion of a protein; and

[0035] Figure 10 shows relative intensity versus mass charge for observed and theoretical isotropic distributions.

DETAILED DESCRIPTION OF THE INVENTION

[0036] In this invention, peptides are labeled in a step separate from the proteolysis step. In one application of the invention, peptides are produced by incubation with the desired serine protease. The peptide products are dehydrated and then incubated with preferably the same type or, in some embodiments a different type of, serine protease under conditions that allow for repeated binding, and disruption of the covalent substrate/enzyme intermediate by O-18 water, sufficient to introduce two isotope labels in greater than 90%, preferably greater than 95% enrichment.

[0037] In this strategy the proteins need not be dehydrated and redissolved. Conditions for the proteolytic reaction can be optimized for proteolysis without concern for the amount of O-18 water. The incorporation of two O-18 labels may also be catalyzed by immobilized proteases, which allows for increased enzyme concentration and more rapid labeling, while controlling contamination by enzyme autolysis.

[0038] A range of times can be utilized for different peptides, reflecting the binding affinities of the enzyme for different peptides. A multiplexed kinetic study is illustrated in Figure 9, in which incubation times for 95% exchange catalyzed by trypsin vary from about 5 minutes to about 90 min. Thus an embodiment of one standard procedure employs incubation times of about 2 hours.

[0039] As the sequencing of many species' genomes is completed, the study of the protein compliment to the genome, or the proteome, has emerged as a dynamic field of research. An approach to characterizing changed states is comparative proteomics, in which the relative amounts of proteins present in two or more samples are compared.

In order to determine the relative amount of the proteins present, a proteolytic method for labeling ^{18}O can be used.

[0040] An exemplary protocol that can be used in conjunction with embodiments of the present invention is described below:

Protocol A

Protein sample preparation

[0041] This protocol describes an optional step to prepare a protein sample for proteolytic labeling.

Materials

Trizma HCl (Sigma)
Trizma Base (Sigma)
Urea (Sigma)
Dithiothreitol (Sigma)
Iodoacetamide (Sigma)
BioSpin 6 Column (BioRad)

[0042] Dissolve protein in 50 mM Tris (pH 8) so that the concentration is 10-200 μg/μL. For the purpose of this recipe the concentration will be 50 μg of protein in 100 μL of base. Tris of pH 8 can be made from Tris base and Tris HCl. Add 0.024 g of urea to make the urea concentration 4 M. (When trypsin is used we can be used to about 4M with a 100 M buffer concentration for Lys-C, 8M urea and for Blu-C and Chymotrypsin 2M urea can be used). Add 10 μL of 100 mM DTT and let rotate at room temperature for 1 hour. The DTT concentration is approximately 10 mM. Add 30 μL of 0.5 M Iodoacetamide and let rotate at room temperature in the dark for 1 hour. The Iodoacetamide concentration is approximately ten times the DTT concentration. Desalt

reduced and alkylated protein sample using a BioSpin 6 Column. One spin column can desalt 75 μL, so two columns are necessary. Sample is recovered in 10 mM Tris, pH 7.4. Pool the recovered samples.

Protocol B

Initial protein digest

[0043] This protocol describes the steps for the digestion of the protein (Figure 9).

Materials

Sequencing grade modified trypsin (Promega)

[0044] Dissolve one vial of 20 μg of trypsin in 20 μL of 50 mM Tris (pH 8). Add 1 μL of dissolved trypsin per 50-100 μg of protein present in sample, or 0.5-1 μL in this case. Remaining trypsin can be stored and used at a later date. Check the pH to ensure that it is approximately 7-8. The pH can be adjusted by adding additional Tris. Incubate at 37°C overnight.

Protocol C

Proteolytic $H_2{}^{18}O$ Labeling

[0045] This protocol outlines the steps for ^{18}O labeling (Figure 9).

Materials

Isotopically enriched $H_2{}^{18}O$ (Isotech, Inc.)
Immobilized trypsin (Applied Biosystems)

[0046] Retrieve digested protein sample from 37°C incubator. Add 10-20 μL of immobilized trypsin to the peptide sample. Immobilized trypsin must be cleaned prior to

use. Wash three times with 5X volume of water, spin on a bench top centrifuge at 8000 rpm for 1 minute and discard the supernatant each time. Dry sample completely in vacuum concentrator. Bring up in 80 μL $H_2{}^{18}O$ and 20 μL acetonitrile. The presence of the organic solvent enhances trypsin efficiency. Store $H_2{}^{18}O$ at 20°C. Rotate at room temperature for 2 to 24 hours. Mix the labeled and unlabeled peptide pools and analyze by mass spectrometry. Samples can be analyzed by nanospray, electrospray, LC-MS, or MALDI and the method is compatible with most commercial mass spectrometers so long as the resolution is sufficient to resolve 4Da.

Protocol D

Interpretation and Protein Quantitation.

[0047] This protocol discusses the interpretation of mass spectral data in order to determine the relative protein quantitation (Figure 10).

Materials

IsoPro 3.0

$$\text{Ratio 1} = \frac{I_4 - \frac{M_4}{M_0} I_0 - \frac{M_2}{M_0}\left(I_2 - \frac{M_2}{M_0} I_0\right) - +\left(I_2 - \frac{M_2}{M_0} I_0\right)}{I_0} \quad \text{(Equation1)}$$

[0048] Ratios of ($^{18}O/^{16}O$) peptides are calculated using Equation 1, where I_0, I_2 and I_4 are the observed peak areas for the monoisotopic peak for the peptides without ^{18}O label, the peak 2 Da higher, and the peak 4 Da higher (double ^{18}O labeled peak), respectively. M_0, M_2 and M_4 are the theoretical peak areas for the monoisotope peak of a peptide with known composition, the peak 2 Da higher and the peak 4 Da higher, respectively. If the peptide sequence is not known an estimate of the $^{18}O;^{16}O$ ratio can

be made by rationing the peak areas of I_4 and I_0. The mixing ratio can be varied from 1:1 and taken into account in the calculations. Values for M_0, M_2 and M_4 can be calculated for each peptide using IsoPro 3.0. The elemental composition and charge state are entered into the software program and the program calculates the theoretical isotopic distribution. The theoretical distribution of the unlabeled and labeled peptides is assumed to be identical.

[0049] In another experiment designed to study the time span required for tryptic peptide products of different sequences, times were determined for complete exchange of a pair of peptides identical except that one is terminated by lysine (YGGFMK) and the other by arginine (YGGFMR). The km values are 4400+-700 microM and 1300+-300 microM, respectively. The kcat values are 2800+-300 per min and 3500+-500 per min, respectively.

[0050] Embodiments of this invention are described in the paper "Dissection of Proteolytic ^{18}O Labeling: Endoprotease-Catalyzed ^{16}O-to-^{18}O Exchange of Truncated Peptide Substrates" (Xudong Yao, Carlos Afonso, and Catherine Fenselau) (Journal of Proteome Research 2003, 2, 147-152), which is hereby incorporated by reference in its entirety.

[0051] Proteolytic labeling in $H_2{}^{18}O$ has been recently revived as a versatile method for proteomics research. To understand the molecular basis of the labeling process, the process has been dissected into two separate events: cleavage of the peptide amide bonds and exchange of the terminal carboxyl oxygens.

[0052] It has been demonstrated that both carboxyl oxygens can be catalytically labeled, independent of the cleavage step. Reaction kinetics of the tryptic ^{16}O-to-^{18}O

exchange of YGGFMR, YGGFMK, and the tryptic digest of apomyoglobin were studied by matrix-assisted laser desorption/ionization Fourier transform ion cyclotron resonance mass spectrometry. A larger K_M for the Lys-peptide (4400 ± 700 μM), when compared to that of the Arg-peptide (K_M 1300 ± 300 μM), was mainly responsible for the slower reaction with YGGFMK (k_{cat}/K_M 0.64 ± 0.14 $\mu M^{-1}min^{-1}$) compared to YGGFMR (k_{cat}/K_M 2.6 ± 0.9 $\mu M^{-1}min^{-1}$).

[0053] Multiplexed kinetic studies showed that endoprotease-catalyzed oxygen exchange is a general phenomenon, allowing homogeneous $^{18}O_2$-coding of a variety of peptides. It was demonstrated for the first time that chymotrypsin $^{18}O_2$-codes peptides during proteolysis.

[0054] Proteolytic ^{18}O labeling can be advantageously decoupled from protein digestion, and endoproteases can be used in a separate step to $^{18}O_2$-code peptides for comparative studies after proteolysis has taken place.

[0055] Tracking the changes in all of the proteins expressed by a cell or tissue in different states provides insight into life processes at the molecular level. Methods enabling the comparison of such changes proteomewide are therefore of great biological and chemical interest. The use of differential stable isotope labeling to code counterpart proteomes in comparison, coupled with mass spectrometric analysis, is one of the most promising approaches. These isotope codes can be incorporated metabolically, chemically, and enzymatically.

[0056] Universal and highly specific proteolytic labeling of proteins in $H_2{}^{18}O$ allows for various applications in proteomic studies, as well as traditional applications in mechanistic studies and sequencing of peptides by mass spectrometry. In comparative

proteomics strategies, endoproteases cleave the first pool of proteins in $H_2{}^{18}O$, and the second pool in $H_2{}^{16}O$.[4]

[0057] The enzymes differentially label the resulting peptides at the carboxyl termini, and generate a mass difference of 4 Da. The isotope ratios of each peptide pair in the combined peptide mixtures are then measured by mass spectrometry. Signal intensities of paired peptides provide relative quantitation of their precursor proteins in the different pools; unpaired peptide signals indicate potential changes in amino acid composition, caused by mutation or co/post-translational modifications.

[0058] Homogeneity and completeness of labeling are beneficial to the successful application of this strategy. Trypsin, Lys-C, and Glu-C can incorporate two ^{18}O atoms into the carboxyl termini of all peptides, except the original protein carboxyl termini. By contrast, some reports have indicated that lysine-terminated peptides do not incorporate two oxygen labels efficiently.[8b-d] Other serine proteases, such as chymotrypsin, have been proposed to incorporate only one ^{18}O during digestion of proteins in $H_2{}^{18}O$.[7]

[0059] In this invention, the process has been mechanistically dissected into two separate events: cleavage of the peptide amide bond and exchange of the terminal carboxyl oxygens (Figure 2).

[0060] Thus, if an endoprotease can catalyze both the hydrolysis and the exchange reactions, then it is able to incorporate two ^{18}O atoms from $H_2{}^{18}O$ during the process, one at each event. The kinetics and mechanisms of the first incorporation do not differ from those of the familiar tryptic cleavage of peptide amide bonds, ignoring isotope effects caused by ^{18}O.

[0061] Therefore, the second incorporation-the exchange reaction-is a key to understanding the molecular basis of the labeling process. Without the exchange reaction, the incorporation would end at one ^{18}O atom.

[0062] Examples have been examined for the catalytic exchange with both trypsin and chymotrypsin.

[0063] Trypsin cleaves peptide amide bonds following arginine (Arg) or lysine (Lys) residues and is the most commonly used protease in proteomics research. The second step, the catalytic exchange, requires trypsin to recognize Arg- and Lys-terminated peptides as substrates. Compared to native peptide substrates, they are truncated and do not have any subsequent P_i' residues. (Amino acid residues from N-terminal to C-terminal of the polypeptide substrate are assigned as ..., P_5, P_4, P_3, P_2, P_1, P_1', P_2', P_3', P_3', P_5', ...,; their respective binding sub-sites in enzyme as ..., S_5, S_4, S_3, S_2, S_1, S_1', S_2', S_3', S_4' S_5'.... The enzymatic cleavage of the amide bond happens between P_1 and P_1'.)

[0064] In embodiments of this invention, YGGFMR, YGGFMK, YGGFLK, and the tryptic digest of apomyoglobin have been used to model Arg-peptides, Lys-peptides, and tryptic peptide mixtures, respectively. The reaction kinetics of the tryptic ^{16}O-to-^{18}O exchange of these substrates were directly studied by matrix-assisted laser desorption/ionization (MALDI) Fourier transform ion cyclotron resonance (FT-ICR) mass spectrometry (MS). The possibility was explored that both carboxylate oxygen atoms could be catalytically exchanged, independent of an amide cleavage step.

[0065] It has been demonstrated that peptide labeling can be decoupled from protein digestion for comparative studies of protein mixtures. In addition, it is demonstrated that chymotrypsin is capable of $^{18}O_2$-coding peptide carboxyl-termini.

Experiment

[0066] Unless otherwise mentioned, all chemicals were reagent grade or better and were purchased commercially and used without further purification. The water ($H_2{}^{16}O$) used in all experiments was purified by a Millipore (Bedford, MA) water filtration system. Isotopically enriched $H_2{}^{18}O$, greater than 95%, was purchased from Isotech, Inc. (Miamisburg, OH). Sequencing grade modified trypsin was purchased from Promega (Madison, WI), and the trypsin concentration was calculated based on weight without active-site titration. Apomyoglobin (horse, muscle), α-chymotrypsin (Bovine pancreas), and peptide VEPIPY were purchased from Sigma (St. Louis, MO). Peptides YGGFM, YGGFMK, YGGFLK, and YGGFMR were purchased from Bachem (Philadelphia, PA).

[0067] Kinetics Measurement of ^{16}O-to-^{18}O Exchange of YGGFMK and YGGFMR Catalyzed by Trypsin. For a typical trypsin-catalyzed ^{16}O-to-^{18}O exchange reaction, peptide YGGFMK (2 mM) was dissolved in $H_2{}^{18}O$ (pH 8.0) containing 18 mM Tris/HCl, 45 mM $CaCl_2$, and the exchange reaction was initialized by addition of trypsin suspended in $H_2{}^{18}O$ to a final concentration of 36 nM (for kinetic study of peptide YGGFMR, final trypsin concentration was typically 4.5 nM). Aliquots of the reaction mixture were taken at different time intervals for the first 10% of reactant conversion. The reaction temperature was maintained at 26 ± 1 °C, and the ionic strength was controlled by NaCl (45 mM). The reaction volume was typically 20 μL. The change of the peak intensity for the peptide was monitored by MALDI-FT-ICR-MS. Substrate concentrations used in the kinetics studies ranged between 0.2 K_M and 5 K_M, and

measurements were duplicated for each substrate concentration. At high concentrations (>10 mM) of YGGFMK, acetonitrile (10% v/v) was used to increase the substrate solubility in reaction buffer. A full course of ^{16}O-to-^{18}O exchange reaction of YGGFMR was also recorded using similar conditions.

[0068] Simultaneous Kinetic Study of ^{16}O-to-^{18}O Exchange of a Peptide Mixture Catalyzed by Trypsin. A mixture of tryptic apomyoglobin peptides ($^{16}O_2$) (100 μM each, obtained by tryptic digestion of apomyglobin in $H_2{}^{16}O$-buffer), YGGFMK($^{16}O_2$) (300 μM), and YGGFLK($^{16}O_2$) (300 μM) was dissolved with DMSO and then diluted with $H_2{}^{18}O$-containg buffer (pH 7.0). The exchange reaction was initialized by the addition of trypsin solution in $H_2{}^{18}O$. The final reaction conditions were 36 nM of trypsin, 18 mM of Tris, 45 mM of $CaCl_2$, 45 mM of NaCl, and 10%v/v of DMSO. Reaction was run at 26 ± 1 °C. At different time intervals, aliquots of the reaction mixture were quenched with 70:30 acetonitrile:water (0.1% TFA) with a dilution ratio of 10 times. Using MALDI-FT-ICR-MS, the changes of the peak intensities for the peptides were monitored for the whole reaction course at their corresponding *m/z* values.

[0069] Chymotryptic Digestion of Apomyoglobin in $H_2{}^{18}O$. An apomyglobin (51 nmol) suspension in $H_2{}^{18}O$ (24 mM Tris/HCl, pH 8.0) was incubated with 200 μg of α-chymotrypsin (bovine pancreas, Sigma) at room temperature. Aliquots were taken at different time intervals for MALDI-FT-ICR analysis.

[0070] MALDI-FT-ICR Mass Spectrometric Analysis. Reaction mixtures in $H_2{}^{18}O$ were diluted with 70:30 acetonitrile:water (0.1% TFA) to obtain a final peptide concentration of about 10 pmol/μl. A 0.6 μL aliquot of this solution was applied on the MALDI stainless steel probe followed by 0.6 μL of a solution of 2,5-dihydroxybezoic acid

(500 mM) in 70:30 acetonitrile:water (0.1% TFA). The mixture was gently stirred during the crystallization. An FT-ICR instrument equipped with a 4.7 T actively shielded magnet (Ionspec Co., Irvine, CA) was used for this study. An external MALDI ion source was differentially pumped. A pulsed (5 ns width) Nd:YAG laser (Minilite, Contimuum, Inc., Santa Clara, CA) was operated at 355 nm (triple of its fundamental frequency). The laser energy was attenuated by adjusting the angle between two Glan-laser prism polarizers (Karl-Lambrecht Co., Chicago, IL). Ions were injected into the ICR cylindrical cell by an RF-only quadrupole ion guide. A 4 ms nitrogen pulse was used to trap the ions and remove their excess of kinetic energy. After a 4 s delay, the ions were excited by a 35 Vb-p RF sweep during 4 ms (29-180 kHz). The image signal was amplified and digitized at an acquisition rate of 2 MHz (512 k data point) resulting in the recording of a 262.1 ms time domain, which gave a typical mass resolution of 10 000. The average of single time domain transients (more than 10 times) was Fourier transformed into the corresponding frequency domain (two zero fills and Blackman apodization). In mass spectra, heights for the monoisotope peaks for peptides, the peaks with masses 2 Da higher, and the peaks with masses 4 Da higher were measured as peak intensities I_0, I_2, and I_4, respectively.

[0071] MALDI-FT-ICR Mass Spectrometric Analysis of ^{16}O-to-^{18}O Exchange Reactions. A relatively short detection time of 262.1 ms for recording the induced current from the coherent motion of ions in the ICR cell gave isotopic distributions close to theoretical ones. Averaging single time domain transients before the Fourier transform resulted in excellent reproducibility (better than ±3%) for the isotope distribution measurement. The concentration of reactant, [*R*], at any reaction time was

calculated according to the following equation: $[R] = [R_0]I_0/\{I_0+(I_2 - M_2I_0/M_0)/2 + [I_4 - M_4I_0/M_0 - M_2/M_0(I_2 - M_2I_0/M_0)]\}$, where $[R_0]$ is the initial concentration of the reactant; I_0, I_2, and I_4 are peak intensities, as assigned for YGGFMK in Figure 3; M_0, M_2, and M_4 are the theoretical relative intensities for the monoisotope peaks for peptides, the peaks with masses 2 Da higher, and the peaks with masses 4 Da higher, which were calculated using Isopro 3.0 (IonSource.com). The residual ^{16}O content in $H_2{}^{18}O$ (less than 5%) was not considered in formulation of the equation. Similar treatment was applied to data for YGGFMR and the apomyoglobin peptide mixture.

[0072] Kinetic Constants for ^{16}O-to-^{18}O Exchange of YGGFMK and YGGFMR Catalyzed by Trypsin. As discussed above, the catalytic exchange was examined using trypsin and peptides that are terminated by K or R. The exchange reactions were monitored by following the diminishing of the I_0 peak and the increase of the I_2 and I_4 peaks (Figure 3) with time, via MALDI-FT-ICR. The contributions to higher masses at time = 0 are made by natural isotope abundance. The time course for the first 10% of the reaction was limited to less than 30 min by adjusting the concentration of trypsin. Partial fragmentation of a peptide during MALDI-FT-ICR analysis did not compromise quantitation. When a fragment signal is strong, following the mass change of the fragment containing the carboxyl terminus gives the identical kinetics as following the intact peptide (data not shown).

[0073] Plots of reactant concentration [R] vs reaction time for the ^{16}O-to-^{18}O exchange reactions were linear, giving slopes twice the initial reaction rates ($2 \times v_i$'s). Double reciprocal plots of $1/v_i$ vs $1/[R_o]$ were linear for both peptides, and the plot for YGGFMR is shown in Figure 4. From the plots, values for k_{cat}'s, K_M's, and k_{cat}/K_M's

were calculated. Peptide YGGFMK has a k_{cat} of 2800 ± 300 min^{-1}, a K_M of 4400 ± 700 μM and a k_{cat}/K_M of 0.64 ± 0.14 $\mu M^{-1}min^{-1}$. Peptide YGGFMR has a k_{cat} of 3500 ± 500 min^{-1}, a K_M of 1300 ± 300 μM and a k_{cat}/K_M of 2.6 ± 0.9 $\mu M^{-1}min^{-1}$. High substrate concentrations (larger than 2 mM) were necessary to make the plots, especially for peptide YGGFMK. Under these concentrations, only minute amounts of coupling products (e.g., YGGFMKYGGFMK) of the peptides were observed, and quantitative correction for this side-reaction was not made. Using pseudo first-order kinetics (initial reactant concentration was smaller than one-tenth of K_M), k_{cat}/K_M's ratios were also measured to be 0.41 ± 0.01 $\mu M^{-1}min^{-1}$ for YGGFMK and 2.3 ± 0.3 $\mu M^{-1}min^{-1}$ for YGGFMR, in good agreement with values from the double reciprocal plots.

[0074] In Figure 5, a full time course of the exchange reaction for peptide YGGFMR shows the decrease of I_0 with time, an increase up to a maximal followed by decrease for I_2 and the increase to a plateau for I_4. I_0 is qualitatively proportional to the concentration of YGGFMR($^{16}O_2$); I_2 to the concentration of YGGFMR($^{16}O^{18}O$), the single incorporation product; I_4 to YGGFMR($^{18}O_2$), the double incorporation product.

[0075] Competitive ^{16}O-to-^{18}O Exchange Reactions of Truncated Peptide Substrates. The tryptic digest of apomyoglobin was mixed with YGGFMK and YGGFLK, and the exchange reaction of the mixture, catalyzed by trypsin, was analyzed by MALDI-FT-ICR. In the mass range of 700 to 2000 Da, apomyoglobin peptides ALELFR (T18), HPGDFGADAQGAMTK (T17), VEADIAGHGQEVLIR (T2), GLSDGEWQQVLNVWGK (T1), and YLEFISDAIIHVLHSK (T16) could be followed to study the reaction kinetics. Other apomyoglobin peptides in the mass range overlapped with fragment ions from heavier peptides. This prevented accurate kinetic analysis. The disappearances of the

$^{16}O_2$-forms of the peptides from the mixture (Figure 6) were fitted to pseudo first-order kinetics, and the observed rate constants were obtained. From these constants, the k_{cat}/K_M values for other peptides (shown in the inset of Figure 6) in the mixture were then calculated, using the k_{cat}/K_M value (0.41 $\mu M^{-1}min^{-1}$) for YFFFMK as an internal standard.

[0076] Exchange Reactions Catalyzed by Chymotrypsin. Chymotryptic proteolysis of apomyoglobin in $H_2{}^{18}O$ was analyzed by MALDI-FT-ICR. In the spectra shown in Figure 8 some of the peptide products of this incubation, RNDIAAKY (CT10), TGHPETLEKF (CT4), and GADAQGAMTKALELF (CT9), can be seen to have monoisotopic masses 4 Da higher than the theoretical ones (MH^+: 949.50, 1158.58 and 1522.76, respectively). This indicates incorporation of two ^{18}O atoms into each peptide. Analysis of the fragments from these ions supported association of the labels with the carboxyl termini. Complete $^{18}O_2$-coding was obtained in 1.5 h for CT10, 3.5 h for CT4, and 20 h for CT9 (see the Experimental section).

[0077] Nature of Trypsin-Catalyzed Oxygen Exchange. A minimal mechanism for the exchange reaction is shown in Figure 7, analogous to the mechanism of serine protease catalysis for cleavage of peptide amide bonds. After formation of the acyl-enzyme adduct between the peptide and the enzyme, a molecule of water hydrolyzes the intermediate and introduces an ^{18}O atom.

[0078] FT-ICR mass spectrometry allows direct and accurate measurement of the initial rates for the exchange reaction, in which isotopic distributions of the reactant ($^{16}O_2$-peptides) and the product ($^{16}O^{18}O$-peptides and $^{18}O_2$-peptides) overlap (Figure 3). During the first 10% of the reaction, where the initial rates were measured, changes in

the intensities of I_2 and I_4 were very small. In addition, the intensity was very low for I_4 relative to that of I_0 (Figure 3).

[0079] The difference in k_{cat} values for YGGFMK (2800 ± 300 min^{-1}) and YGGFMR (3500 ± 500 min^{-1}) was minimal. Weaker substrate binding for the Lys-peptide (K_M 4400 ± 700 μM) compared to the Arg-peptide (K_M 1300 ± 300 μM) was mainly responsible for the slower reaction with YGGFMK (k_{cat}/K_M 0.64 ± 0.14 $\mu M^{-1}min^{-1}$), compared to YGGFMR (k_{cat}/K_M 2.6 ± 0.9 $\mu M^{-1}min^{-1}$). Linearity of double reciprocal plots of $1/v_i$ vs $1/[R_0]$ for both the peptides (Figure 4 for peptide YGGFMR) suggests that substrate/product inhibition was not significant under the experimental conditions. Equilibrium constants are compatible with K_M's for chymotrypsin-catalyzed oxygen exchange reactions. Analogous to this example, we assumed that K_M values for the trypsin reactions gave reasonable estimations for substrate binding. Lacking all P_i'-S_i' interactions, the affinity of trypsin for truncated substrates is intrinsically lower than that for larger polypeptide substrates (with both P_i and P_i' sites). Hydrolysis of a native polypeptide substrate by trypsin has been reported to occur with a higher k_{cat}/K_M (18 $\mu M^{-1}min^{-1}$), lower K_M (140 μM) and similar k_{cat} (2500 min^{-1}). For very short, truncated peptide substrates, some P_i-S_i interactions are missing in addition to nonexisting P_i'-S_i' interactions; and the exchange reaction was even slower (data not shown). Mass spectra of such short peptides do not typically provide unambiguous identification of their precursor proteins.

[0080] The fact that trypsin catalyzes the exchange reaction of both carboxyl oxygens of YGGFMK with oxygen atoms from water proves that the protease is able to $^{18}O_2$-code peptides that end in lysine. If the isotope effects from ^{18}O are omitted, then

the rates of both incorporations are identical. Previous observations that only a single ^{18}O could be incorporated into Lys-terminated peptides during proteolysis of proteins in $H_2{}^{18}O$[8b-d] appear to have resulted from nonideal reaction conditions, e.g., insufficient enzyme concentration or insufficient reaction time. In the case where the labeling process starts with the digestion of proteins, diffusion of proteins or incomplete cleavage fragments to the enzyme active site could present a rate-limiting step.

[0081] For the process of $^{18}O_2$-coding of Arg- and Lys-terminated peptides through two exchange reactions, a sequential mechanism (Figure 7) is supported by the kinetics shown in Figure 5. Trypsin-catalyzed exchange reactions are substrate-specific; only peptides with Arg or Lys in the P_1 position can undergo exchanges catalyzed by trypsin. As a control experiment, a mixture of VEPIPY (710 μM) and YGGFM (500 μM) was incubated with 200 mM Tris and 2 μg trypsin in $H_2{}^{18}O$ (pH 8.5) at room-temperature overnight. No detectable exchange of oxygens, even the first oxygen exchange, was observed. Thus, there was not significant catalysis from either the enzyme or the buffer.

[0082] Kinetic Study of Simultaneous Exchange Reactions of Peptide Mixtures. The time required for complete ^{18}O exchange of the mixture multiplexed in this experiment was less than 2.5 h, and k_{cat}/K_M ratios varied from 0.089 ± 0.006 $\mu M^{-1}min^{-1}$ for YGGFLK to 1.1 ± 0.2 $\mu^{-1}min^{-1}$ for ALELFR (T18 of apomyoglobin), which further demonstrates the general applicability of the exchange reaction. The range of k_{cat}/K_M ratios for a peptide mixture can vary by several orders of magnitude, as estimated using data from a study on trypsin subsites. Furthermore, this study provides an example of the simultaneous study of multiple reactions, using the capabilities of MALDI-FT-ICR: large dynamic

range, ultrahigh resolution, and the fact that MALDI predominantly generates singly charged ions.

[0083] There are often peptides (e.g., hydrophobic peptides) in a mixture that are difficult to dissolve directly in aqueous buffers. When a peptide was not fully dissolved in solution, two populations of the peptide ($^{16}O_2$-peptide and $^{18}O_2$-peptide, 4 Da difference in mass) were observed in mass spectrometric analysis. This suggests that after reaching steady-state (all dissolved molecules of a peptide having exchanged their carboxyl terminal oxygen atoms), solubilization of the peptide is the rate-limiting step in the labeling reaction. Before reconstitution of peptide mixtures with $H_2{}^{18}O$ buffers for the exchange reaction, dissolution of the entire mixture with a strong solvent such as DMSO is beneficial to successful homogeneous labeling.

[0084] The use of fluorescent peptide libraries has been recently reported to provide high-throughput profiling of protease specificities for peptides with different S_i and S_i' sites. An alternative strategy can be provided by ^{16}O-to-^{18}O exchange reactions, which eliminates the need for conjugation of fluorescent labels. If sequences of Arg-terminated peptides are systematically changed and synthesized, then trypsin-catalyzed ^{16}O-to-^{18}O exchange reactions of such peptide libraries could be used to profile trypsin specificities for S_i sites. Analogous experiments can be designed for other proteases. The capability of MALDI-FT-ICR to monitor multiple reactions simultaneously increases the throughput of this method.

[0085] $^{18}O_2$-Coding of Peptides by Chymotrypsin. It is demonstrated here that chymotrypsin can $^{18}O_2$-code peptides during proteolysis. The different times for complete $^{18}O_2$-coding appear to reflect recognition preferences of the enzyme active

site for the carboxyl terminal sequences of the peptides. Chymotrypsin and other serine endoproteases are known to catalyze oxygen exchange of carboxylic acids. However, chymotrypsin was believed to label peptide termini with only one ^{18}O during proteolysis in $H_2{}^{18}O$. The mechanistic dissection of the proteolytic labeling process reveals this observation to be experimentally dependent, not reflecting the intrinsic catalytic capability of chymotrypsin. Compared to trypsin, chymotrypsin catalysis of oxygen exchange reactions was less efficient, but with sufficient enzyme concentration, as can be achieved by immobilized forms of the enzyme, the exchange could be completed on a practical time scale. Recognizing large hydrophobic residues as P_1 sites, this enzyme could be useful in comparative proteomic studies of hydrophobic protein mixtures.

[0086] It is demonstrated above that various endoproteases catalyze the oxygen exchange reaction of appropriate truncated peptides, independent of the cleavage of peptide amide bonds in the precursor proteins. (On the basis of mechanistic similarity, all of the serine proteases, if not other classes of proteases, that cleave C-side of P_1 residues should have the same capability. Trypsin, chymotrypsin, Lys-C, and Glu-C are among these proteases.) Therefore, proteolytic ^{18}O labeling can be decoupled from protein digestion and to use endoproteases to $^{18}O_2$-code peptides after proteolysis.

[0087] In this strategy, the two protein mixtures being compared are digested separately in $H_2{}^{16}O$ under optimal conditions and dried. Each resulting peptide mixture is then reconstituted with $H_2{}^{16}O$ or $H_2{}^{18}O$ for enzyme-catalyzed differential labeling with stable isotopes of oxygen. After combination of the two pools of labeled peptides, concurrent separation and mass spectrometric characterization will provide accurate comparisons of changes in expression and modification of proteins in the counterpart

mixtures. It is worthwhile to discuss that although MALDI-FT-ICR was used in quantitative study of oxygen exchange kinetics, other mass spectrometers can serve well for peptide/protein quantification enabled by $^{18}O/^{16}O$-labeling. In fact, the resolution for FT-ICR was set to be ca. 10 000, which is in the resolution range for mass spectrometers with configurations such as quadrupole-time-of-flight, reflectron-time-of-flight or triple quadrupole. When doubly or triply charged ions of differentially labeled peptides have small mass difference as 3 or 4 Da, use of low-resolution mass spectrometers such as ion traps can be difficult; a new mass spectrometric method has been proposed to address this issue.

[0088] The decoupling has the advantages that proteins can be kept in solution prior to digestion, that proteolytic peptide products can be labeled with a limited volume of $H_2{}^{18}O$, and that digestion and labeling conditions can be optimized separately. All of these features are of real interest in investigations of complex protein mixtures, and they allow more effective and versatile applications of enzymatic labeling. This decoupling strategy can be used, for example, to study changes of proteins in human cancer cells upon acquisition of drug-resistance.

[0089] It has been shown that chymotrypsin can catalyze full exchange of two O-18 atoms into peptides recovered from a separate chymotrypsin proteolysis step, as indicated in Figure 10. The reaction time is slower for catalysis by chymotryptic than for tryptic catalysis, however, this exchange based on binding and on the mechanism of the enzyme.

[0090] A method in accordance with embodiments of this invention has been utilized for investigating doxorubicin resistance in MCF-7 Breast Cancer cells (see the paper

"Investigation of Doxorubicin Resistance in MCF-7 Breast Cancer Cells Using Shot-Gun Comparative Proteomics with Proteolytic 18O Labeling, Brown and Fensalau, Journal of Proteome Research, Received October 6, 2003). The subject matter of that paper is incorporated herein in its entirety.) In that paper, the decoupling of the digestion and labeling steps eliminated the need to dry samples at the protein level. To further optimize the ^{18}O labeling strategy, the use of immobilized enzymes was implemented. This strategy was optimized for the application to complex protein mixtures. The use of an immobilized enzyme allowed for a high enzyme concentration with minimal autolysis products. This extremely high enzyme concentration allowed for a shorter reaction time and more complete incorporation.